## REMARKS

The indication of claims 5, 10, and 12 containing allowable subject matter is noted.

Claims 1 and 12 have been amended to overcome the rejection thereof based on 35 USC 112, paragraph 2.

Claim 8 has been cancelled and replaced by claim 13 to overcome the rejection thereof based on 35 USC 112, paragraph 2 and the rejection based on 35 USC 101.

Amended claims 1-4, 6, 7 and 9 and new claim 13 are not rendered unpatentable under 35 USC 103(a) as a result of Van den Akker (US 6,415,250) in view of De Campos (US 6,272,456), the combination of references previously relied on to reject claims 1-4 and 6-9.

Applicant repeats the remarks filed on March 18, 2009, relating to the final office action, in response to rejection the examiner's arguments in paragraphs 3 and 9 of the office action at page 3 relating to the rejection of claims 1-4 and 6-9.

In addition, paragraph [0014] of applicant's published application indicates the language identification system of Van den Akker is limited to a single category of first category strings, such as suffixes or prefixes or infixes, in a word (column 8, lines 5-12, figure 5A). Column 20, lines 36-43, of Van den Akker states prefixes and suffixes of words can be extracted. This does not mean one or plural prefixes and one or plural suffixes are extracted from one word. Instead, it means one suffix or one prefix is extracted from one word, i.e., one prefix can be extracted from a word and one suffix can be extracted from another word. Van den Akker, at column 8, line 63 to column 9, line 3, and in claims 1, 3, 12 and 15 refers to suffixes and word endings. These portions of the references do not state that a combination of all the portions of one word can be extracted. Therefore, Van den Akker does not include the requirements of amended independent claim 1 or new independent claim 13 to analyze words extracted from digital text, to thereby construct for each extracted word all the character strings contained in said extracted word, including all the prefixes, suffixes and infixes in said extracted word, with overlap and different lengths lying between one character and the number of characters in said extracted word. Van

den Akker analyzes only one character string per extracted word relative to the corpus of a language, irrespective of the position of the character string in the word (and therefore, of the length of the character string if the character string is a suffix or prefix).

The significance of the foregoing differences between applicant's claims 1 and 13 and Van den Akker can be seen by comparing certain parts of the Van den Akker specification and applicant's specification.

In Van den Akker [column 11, line 65-column 12, line 18], a "suffix of the each of the parsed words 403" in the suffix extractor 404 "is determined to have a predetermined number of characters at the end of each parsed word," and "the predetermined number of characters in the parsed words 403 extracted by the suffix extractor 404 is three characters," and "the source language of an unknown text 301" is identified "by analyzing the last three characters of the parsed words 403," and "for example, four characters at the end of the word may be used to capture the suffix." In the same manner, the predetermined number of characters, three or four, defines the length of a prefix in the parsed word, as indicated by column 8, line 58 to column 9, line 3, which states: "word portions that contains the prefix may be used"; Fig. 2C. Obviously, when the parsed word contains fewer than the predetermined number of characters, for example equal to three, the extracted suffix or the extracted word portion with three or two or one characters is the word itself (column 12, lines 27-32).

In De Campos (column 10, lines 46-65), as a three-letter windows is slid over a particular training document as "ABCABCKLM", the "ABC" letter sequence would appear twice and the "KLM" letter sequence would appear only once. However, De Campos has no disclosure of all the other letter sequences having more or less than three letters and overlapping, such as "BCABC", "CABCKL", "ABCABC" and "AB."

Paragraph [0032] of applicant's published application states, "the first three directories PRq, SUq and INq relate to morphemes, syllables and short character strings CH of from one to six characters, for example" and more particularly relate respectively to prefixes, suffixes and infixes in the predetermined languages. An infix is defined as a character string that is included between the first character, i.e., the start, of a word, and the last character, i.e., the end of the word, as e.g. "ou" or "oi" [paragraph 0036 of

applicant's published application]. In other examples, such as found in Van den Akker, at column 8, lines 53-57 (see also figures 2B-2C), applicant's directories include the suffix "ing" of the word "smashing", but also includes the infixes "mash" and "ash" of the word "smashing." The applicant's directories also include the suffix "ment" of the word "development," and the infixes "velo", "ve", and "lo" of the word "development".

As a result of the above remarks, the word analyzing in the claimed device of claim 1 and method of claim 13 constructs all the character strings contained in an extracted word and having lengths lying between one character and the number of characters in said extracted word, such as the prefix "de", the suffix "ment", and the other character chains "eve", "evelo", "lopm", "opme", etc., having infixes "velo", "ve", and "lo", included in the extracted word "development" and thus including the partially overlapping chains as stated at paragraph [0057] of applicant's published application by: "In the latter variant, the character strings CH contained in the extracted word MT and found in the directories PRq, SUq and INq may partially overlap, in contrast to the n-grams of the approach disclosed in US patent No. 6,292,772 B1 already commented on. For example, if the processed word MT is the French word "aiment", the character strings "ment" and "ent" placed in the pseudo-suffix directory SUq overlap in the processed word. Another example is the overlap of infix "oi" and the pseudo-suffix "is" of the processed word "vois."

Therefore, applicant repeats his position that Van den Akker fails to disclose the function of the claimed analyzing means or step and more particularly the means for or step of analyzing words extracted from said digital text to thereby construct for each extracted word all the character strings, as set forth in paragraph 0049 of applicant's published application which indicates analyzer 6 analyzes each extracted word MT to construct all the character strings CH included in the extracted word MT, including all the prefixes and suffixes in the extracted word, as well as the character chains included between the start and end of the extracted word and overlapping (e.g. infixes and pseudo-infixes), and having lengths lying between one character and the number of characters in said extracted word.

Van den Akker also fails to disclose the requirements of claims 1 and 13 relating to (1) the pre-stored first character strings, including all the prefixes and suffixes of words of a

plurality of predetermined languages and character chains included between the start and end of said words and overlapping (including infixes and pseudo-infix infixes), that occur frequently anywhere respectively in said words of said predetermined languages, and (2) the pre-stored second character strings of different lengths (including prefixes and suffixes) in said words of said plurality of predetermined languages and character chains included between the start and end of said words and overlapping (e.g. infixes and pseudo-infix infixes), that are atypical anywhere respectively in words of said predetermined languages.

Applicant also disagrees with the contention in the office action that Van den Akker bases the score of a character chain extracted from a parsed word on the position of the first character chain in the extracted word of a parsed word. For example, the first character chain of a Van den Akker parsed word is a suffix or a prefix and always has the same position in the extracted words, i.e., the last position in all the parsed words, ended by said suffix, counted from the last character in the parsed words (or for a prefix, the first position in all the parsed words starting with said prefix, counted from the first character in the parsed words).

The position of a character chain in a parsed word is not disclosed by Van den Akker in the paragraph in column 9, lines 18-41, cited in the office action. This relied on portion of Van den Akker only relates to the frequency of a character chain in words of a language corpus. A selected word portion (character chain), such as a suffix, also called a "word ending" (or a prefix) "was extracted from the corresponding language corpus 309 along with a frequency value indicative of the number of times the selected word portion was found within the corresponding language corpus 309. These frequencies are subsequently normalized to a common corpus size, serving as relative frequencies of each of the selected word portion in each language; column 9, lines 34-41 and the frequency lists in column 13, lines 33-42. The frequency value is indicative of the number of times the selected word portion was found within the corpus and does not take into consideration the position of the character string in the parsed word. For example, one suffix and one infix having the same frequency value can have different positions, and two suffixes or prefixes having the same length can have different frequencies.

Applicant's claims 1 and 13 indicate the position of a character chain in an extracted word defines a first coefficient. This limitation differs from (1) the frequency value of a selected word portion disclosed by Van den Akker and (2) the frequency of a character chain in an extracted word, as defined in applicant's claims 3 and 5, and disclosed by paragraph [0059] of applicant's published application. For example, the coefficients PO, FR and LON respectively depend on (1) the position of the character string CH in the extracted word MT, (2) the frequency of the character string CH in the determined language, Lq, and (3) the length of the character string CH. For example in the extracted words "development", "abutment" and "government", the suffix "ment" always has the last position.

Van den Akker does not consider any distinction relating to the position of the character chain "ment" in these three words. In contrast, applicant assigns three different coefficients depending on the positions 8, 4, 3, 6 and 5 of the character chain "ment," counted from the beginning of the words "development", "alimented", "lamentable", "incremental" and "rudimental," or the positions 4, 6, 8, 6 and 6 of the character chain "ment" counted from the ends in the five words respectively. In a corpus including the words "development", "abutment", "government", "alimented", "lamentable", "incremental" and "rudimental", the frequency of the character chain "ment" is seven and different from the values of the positions.

Therefore, Van den Akker fails to include the requirements of claims 1 and 13 for a score that is calculated by adding to the score a first coefficient whenever a prestored first character string of said one determined language is found in said extracted word, wherein said first coefficient depends on the position of the found prestored first character string of said one determined language in said extracted word.

Van den Akker, at column 13, line 62 to column 14, line 24, effectively discloses that the probability, i.e., the normalized frequency value, of a word portion is negative to indicate a strong likelihood that the language of the input text is not the corresponding language. However, the Van den Akker word portion can be a suffix, while a second coefficient in claims 1 and 13 is associated with a found prestored second character string which can be a prefix, a suffix or any word chain included between the start and end of

words of the predetermined languages and overlapping (including infixes and pseudo-infix infixes), as disclosed by 0042-0044 of applicant's published application.

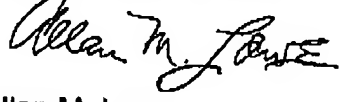
Because of the foregoing differences between claims 1 and 13 and Van den Akker, that are not considered in the office action, there is no need to discuss de Campos that the examiner alleges concerns the second coefficient associated with a second character string that depends on a predetermined language. This is particularly the case because neither Van den Akker nor de Campos discloses the requirements of claims 1 and 13 relating to a first coefficient that depends on the position of a character string found in an extracted word.

Because claims 1 and 13 are allowable over Van den Akker in view of de Campos, dependent claims 2-4, 6, 7 and 9 are also allowable.

To the extent necessary, a petition for an extension of time under 37 C.F.R. 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account 07-1337 and please credit any excess fees to such deposit account.

Respectfully submitted,

**LOWE HAUPTMAN HAM & BERNER, LLP**



Allan M. Lowe
Registration No. 19,641

1700 Diagonal Road, Suite 300
Alexandria, Virginia 22314
(703) 684-1111
(703) 518-5499 Facsimile
**Date: September 22, 2009**
**AML/cjf**